

Classification of Underwater Objects Via Impulse Excitation

J. Tory Cobb, Rodolfo Arrieta, and Charles Bernstein
Naval Surface Warfare Center
Panama City
Panama City, Florida 32407

Abstract—A technique for classifying objects based on modeling the transient characteristics of their impulse response is developed and tested. A set of targets identical in geometry and differing in shell and filler material were constructed. The targets were manually struck exciting an impulse response which was sampled and recorded. The impulse response of each target was decomposed via windowed short-time Fourier transform into a set of feature vectors. The feature vectors were quantized via the LBG VQ algorithm, and the sets of quantized vectors were used to estimate the parameters of a discrete-output hidden Markov model (HMM) for each class of object. A blind test set was evaluated against the trained HMMs and the results are presented along with a discussion of the generalization ability of the individual classifiers.

I. INTRODUCTION

The acoustic impulse response of a geometrically-complicated object is composed of many sinusoidal modes, each with different damping coefficients and fundamental frequencies. This impulse response can be difficult to predict if the object's dimensions cannot be precisely predicted, the construction materials are not homogeneous, or the impinging location of the excitation pulse is highly variable. However, repeatable time-evolving frequency features can be observed in the impulse response even when the underlying generating physical phenomena are not well understood.

Acoustic signal pattern classification techniques such as those used in speech recognition are a natural fit for classifying underwater objects using these time-evolving frequency features. Among these pattern classification methods, hidden Markov model (HMM) classifiers have been used to successfully classify speech signals for many years [1], [2]. In many speech applications, the speech signal is modeled as a concatenation of primitive speech elements called phonemes and the classification task is to estimate the transition probabilities of these underlying primitives and encode them as states in the HMM.

The approach described in this paper similarly treats the time-frequency decomposition of the acoustic impulse response as a time-evolving set of emissions from an underlying hidden state sequence that is unique to each class of object. This approach has been successfully used in other underwater acoustic signal classification tasks in discriminating tonal signals from chirp and continuous-wave pulses [3]. Additionally, rather than use a standard time-frequency decomposition of the acoustic return, others have used different basis functions

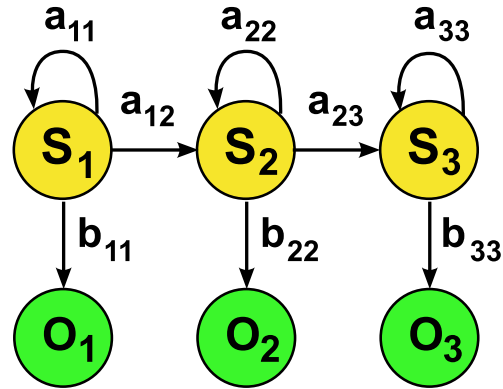


Fig. 1. Graphical representation of a left-right hidden Markov model. The nodes labeled with variables S_1 , S_2 , and S_3 represent the underlying state sequence. The nodes labeled O_1 , O_2 , and O_3 represent the possible output states.

matched to predicted scattering wave physics as feature inputs to HMM classifiers [4], [5].

The following sections briefly introduce the reader to the HMM classifier and describe the acoustic impulse response experiment, classifier training and evaluation, and test results.

II. HIDDEN MARKOV MODEL CLASSIFIER

A discrete-output hidden Markov model is defined by the three parameters: \mathbf{A} the state transition matrix, \mathbf{B} the state emission probability matrix, and π the vector of initial state probabilities. As the term Markov implies, the conditional probability of transition from the current state s_1 to state s_2 , $P(S = s_2 | s_1)$ depends solely on the current state s_1 . These discrete transition probabilities populate the state transition matrix \mathbf{A} where the entry a_{ij} is the conditional probability $P(S = s_j | s_i)$. If the labels of each state are known by observing the data, \mathbf{A} can be estimated directly. However, usually the labels of each state are unknown or "hidden" and the entries of \mathbf{A} are estimated by iterative Expectation-Maximization techniques such as the Baum-Welch algorithm [1]. The state emissions probabilities can be modeled as either discrete or continuous random variables conditioned on the underlying state. In the discrete case, \mathbf{B} is a matrix where entry b_{ij} is the conditional probability of emitting observation o_i given state s_j or $P(o_i | S = s_j)$.

Figure 1 illustrates the concepts described in the previous paragraph. The nodes labeled S_1 , S_2 , and S_3 in the top of

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 01 SEP 2006		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Classification of Underwater Objects Via Impulse Excitation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Surface Warfare Center Panama City Panama City, Florida 32407				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM002006. Proceedings of the MTS/IEEE OCEANS 2006 Boston Conference and Exhibition Held in Boston, Massachusetts on September 15-21, 2006, The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 5	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

the figure represent the underlying state sequence. The state transition probabilities are denoted by a_{11} , a_{12} , a_{22} , a_{23} , and a_{33} . The state emission probabilities for the observations \mathbf{o}_1 , \mathbf{o}_2 , and \mathbf{o}_3 are denoted b_{11} , b_{22} , and b_{33} respectively. (Note that the emission probabilities are depicted only for the given observation sequence \mathbf{o}_1 , \mathbf{o}_2 , and \mathbf{o}_3 , the other emission probabilities are not shown. Had the output sequence been ordered \mathbf{o}_2 , \mathbf{o}_3 , \mathbf{o}_1 , the emission probabilities would be labeled b_{12} , b_{23} , and b_{31} respectively.)

Additionally, this particular figure illustrates a subclass of HMMs that are called left-right HMMs. In a left-right HMM a current state can only transition to another state with an index greater than or equal to the current index. The left-right constraint is a realistic assumption when previous states cannot be revisited due to a phonetic or physical limitation. For example, when modeling the utterance of the word "cat", it is a reasonable assumption that the state that emits the primitive associated with the hard /k/ sound cannot follow the state that emits the primitive associated with the /ă/ sound. This assumption reduces the number of entries in the \mathbf{A} matrix, thus simplifying the estimation process.

In a typical speech processing application, many utterances of a word or sequence of words are digitally sampled and recorded. The time series for each utterance is in turn parsed into a series of representative features such as spectral or linear predictive coding coefficients to form a sequence of N observation vectors $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_N\}$. To train the HMM, maximum likelihood estimates of the parameters \mathbf{A} , \mathbf{B} , and π are found that maximize the product of evaluation probabilities, $\prod_{i=1}^M P(\mathbf{O}_i | \mathbf{A}, \mathbf{B}, \pi)$, across the ensemble of M observation sequences, where $P(\mathbf{O} | \mathbf{A}, \mathbf{B}, \pi)$ is the probability an observation sequence \mathbf{O} is generated by the HMM with parameters \mathbf{A} , \mathbf{B} , and π . Once the appropriate number of HMMs have been trained, new observation sequences can be evaluated for classification. When evaluating an unknown observation sequence \mathbf{O} against a set of k trained HMMs, \mathbf{O} is associated with the HMM that evaluates to the largest probability $P(\mathbf{O} | \mathbf{A}_k, \mathbf{B}_k, \pi_k) \forall k$.

This section briefly described the key HMM parameters and equations that are directly related to the experiment in this paper. For an in-depth review of HMM parameter estimation via the Baum-Welch algorithm and the algorithmic steps to evaluating the expression $P(\mathbf{O} | \mathbf{A}, \mathbf{B}, \pi)$ the reader is directed to Rabiner's excellent tutorial and Rabiner's and Juang's textbook listed in the bibliography [1], [2].

III. ACOUSTIC IMPULSE CLASSIFICATION

For the experiment, a set of 9 geometrically identical objects were constructed by placing end caps on hollow cylinders made of three different shell materials which were filled to capacity with three different filler materials. The different object classes will hereafter be labeled according to a two number system, where Object ij means the object with shell material Type i and filler material Type j . The shell and filler materials are numbered from 1 to 3 in order of increasing density. To generate a data set for training and evaluating each object's



Fig. 2. Block diagram of Acoustic Impulse Classification. A time series is preprocessed, decomposed via windowed STFT and quantized, then evaluated against several trained HMMs. The class label of the HMM that evaluates to the highest probability is assigned to the input observation sequence.

HMM, each object is struck 15 times near its lengthwise midpoint and the acoustic response is sampled and recorded following each strike. Each time series is then preprocessed and parsed into a set of quantized feature vectors using time-frequency decomposition followed by vector quantization. For the task of training each HMM, a subset of 8 time series are used to estimate the HMM parameters for each object class. For the case of object classification, an unknown test pattern is associated with the object whose HMM evaluates to the largest probability over the set of all possible HMMs. Figure 2 depicts the block diagram of the evaluation process described above.

A. Preprocessing

Each time series is prepared for time-frequency decomposition and vector quantization by three preprocessing steps. Since most of the discriminating target information was found to be in lower frequency bands, each time series is first filtered and downsampled to simplify computation and storage requirements. Second, each down-sampled time series is normalized to unit energy. The striking energy for each generated time series is highly variable, thus making total energy of the acoustic impulse response an unreliable discriminating feature. Finally, the "dead zone" that exists before the strike occurs is removed by ignoring samples until the total integrated energy rises above a certain threshold.

B. Time-Frequency Feature Vectors

As shown in Figure 3, the time-frequency plot of an object's acoustic impulse response is composed of several clustered distinct events with varying frequency and time characteristics. This plot gives insight into why one would use a time-evolving stochastic model such as an HMM to discriminate between such signals. These time-frequency events are unique to each class of object and can be modeled directly by the state-sequence transition and state emission matrices of the HMM.

To put the time-frequency information into a format that is compatible with discrete-output HMM training and evaluation, a windowed short-time Fourier transform (STFT) followed by vector quantization is performed on each time series. Following preprocessing, the time series data is parsed into 40 overlapping frames of 100 time samples. Each frame is then windowed by a Gaussian window to reduce spectral leakage and the DFT magnitude of each frame is recorded. Additionally, since the time series signal is real, only the first 50 DFT magnitude coefficients are retained since the remaining 50 are redundant. In more compact notation, the observation sequence \mathbf{O} is encoded as a set of 40 vectors of DFT magnitude $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_{40}\}$ where 40 is the number of STFT frames.

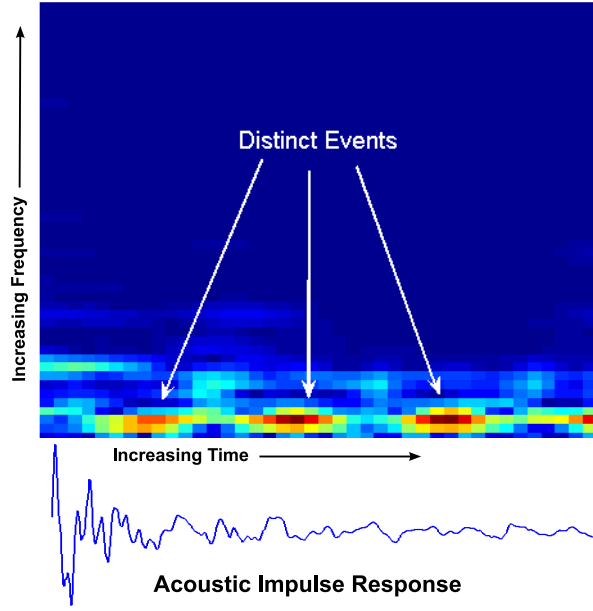


Fig. 3. Time-Frequency decomposition of an acoustic impulse response. Note the several distinct events identified by the white arrow.

In a discrete-output HMM, the emission of the underlying states has a discrete probability distribution, meaning each vector in the sequence \mathbf{O} must be quantized into a representative symbol that relates the underlying DFT magnitude vector to a possible output of the hidden state. To accommodate this requirement the LBG vector quantization (VQ) algorithm was used to quantize the observation feature vector space into 32 discrete states for all possible targets using the 2880 ($9 \text{ objects} \times 8 \frac{\text{training sequences}}{\text{object}} \times 40 \frac{\text{vectors}}{\text{training sequence}}$) feature vectors generated from the training sequences of time series data [6].

C. HMM Training

The quantized time series training sequences of feature vectors are separated according to class and used to train class-specific HMMs. Prior to parameter estimation, the entries of \mathbf{A} and \mathbf{B} are initialized to random values and the π vector is assigned the value $\{1, 0, \dots, 0\}$ since it is a left-right HMM and therefore must begin at the first state. What remains is to specify the number of hidden states in the state sequence and thus the sizes of the \mathbf{A} and \mathbf{B} matrix prior to maximum likelihood estimation using the Baum-Welch algorithm. An empirical method was used to determine the number hidden states by training each HMM using an increasing number of hidden states until a training objective function increased to a reasonable value and remained steady. The training objective function for an HMM of the j -th class over a series of N observations is defined as

$$\Phi(j) \triangleq \frac{1}{N} \sum_{i=1}^N \left(\argmin_k \left[\ln P(\mathbf{O}_j^{(i)} | \lambda_j) - \ln P(\mathbf{O}_j^{(i)} | \lambda_k) \right] \quad \forall k \neq j \right)$$

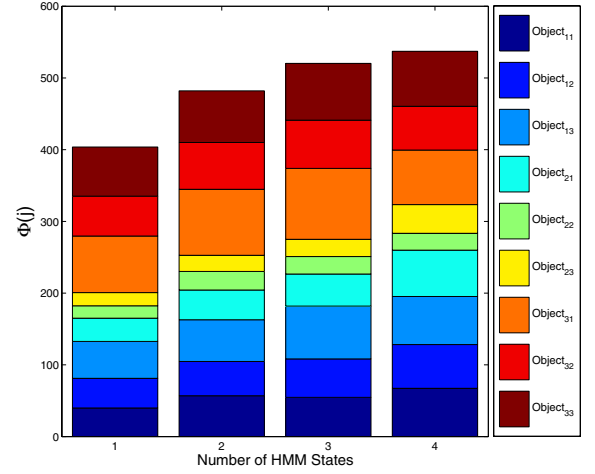


Fig. 4. Stacked bar graph of the training object function values for a given number of HMM hidden states. The size of each element in the stack indicates how well the trained HMM discriminates between objects.

where $\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$. This function estimates the mean value of a difference of log probabilities between all training sequences associated with Object j evaluated on Object j 's HMM and the HMM that gives the next largest output. This function is more useful than strictly relying on the training misclassification error, especially when the misclassification error is extremely low and there a small number of training sequences. The larger the value, the better the trained HMM is at correctly classifying the training sequences. A very small or negative value of $\Phi(j)$ indicates the trained HMM has a high misclassification rate with the training sequences.

The vertically stacked bar graph in Figure 4 depicts the training objective function values for a given number of HMM hidden states in ascending class order starting at object class 11 at the bottom of the stack and ending with object class 33 at the top of the stack. As the number of hidden states increases from 1 to 4 the overall training objective function contributions trend upward. Different numbers of hidden states were chosen to set HMM training parameters for each object class based on the lowest number of states for which the value of $\Phi(j)$ failed to increase appreciably. Based on this criterion Objects 11, 21, 22, and 23 were assigned 2 hidden states, and Objects 12, 13, 31, 32, and 33 were assigned 3 hidden states. This approach favors simpler models for each class, which is crucial when estimating entries in the state transition matrix \mathbf{A} in a scenario with so few training examples. Although the plot shows objective function values for up to 4 hidden states, the $\Phi(j)$ values in column 4 become less reliable due to the extra training sequences needed to estimate the \mathbf{A} matrix. No 4-hidden-state models were chosen for this reason.

While not mentioned above, the use of 1 hidden state is a special case of the HMM where predictably, $\mathbf{A} = \mathbf{I}$, and the evaluation probability of an observation sequence is solely dependent upon its emission probabilities defined by

B. The classification task with single state HMMs is similar to other single-input single-output classifiers such as those implemented by neural networks or Bayesian classifiers. As is shown in Figure 4, classification performance on the training examples with a single state HMM is actually quite good. This indicates the different object classes' observation sequences do not intersect over many quantized feature vectors.

D. HMM Evaluation

After each HMM is trained by estimating the parameters **A** and **B** from the training data, the classifier is ready to receive unknown test patterns. Referring again to Figure 2, the time series data is preprocessed by downsampling and then decomposed via a windowed STFT. The STFT frames are quantized by assigning them the discrete symbol that is closest in Euclidean distance to the VQ centers determined in Subsection III-B to form the observation sequence $\mathbf{O} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{40}\}$, where \mathbf{v}_i is the symbol associated with the i -th VQ output cluster. This observation sequence is in turn used to evaluate each trained HMM, and the object class label k that evaluates to the largest probability in the evaluation equation $P(\mathbf{O}|\mathbf{A}_k, \mathbf{B}_k, \pi_k)$ is assigned to the unknown pattern.

IV. RESULTS

A set of 7 blind test examples were evaluated against the trained HMMs. The tabulated results of the object classifications are presented as fractions of correct classifications to total test sequences in Table I. While the results were good for the test sequences listed in the table below, with so few test examples it is difficult to draw meaningful conclusions about the classifiers' generalization abilities. As another measure of performance, the calculation of the objective function $\Phi(j)$ was repeated for the test patterns to create the stacked bar graph shown in Figure 5. The graph is plotted using the same axes of Figure 4 to give some intuition of how robust each trained HMM is to novel test patterns. As can be expected, the margin of the evaluation probabilities between correct and incorrect classes has shrunk considerably across all class labels. Additionally, in the 4-state column the performance behavior of some HMMs becomes erratic, suggesting overtraining or lack of training data to sufficiently estimate the HMM parameters. By cross-comparing the classification results in one cell of Table I to the size of the corresponding bar graph element, one can gain a sense of how well the correct model is separated from the other models for a given set of test sequences. For example, Object 11 was correctly classified

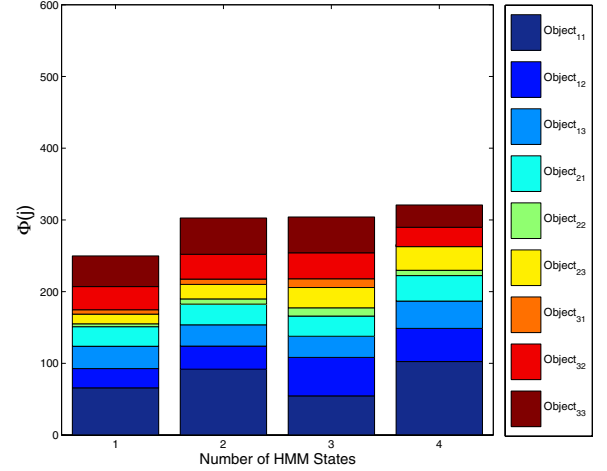


Fig. 5. Stacked bar graph of the test object function values for a given number of HMM hidden states. The size of each element in the stack indicates how well the trained HMM discriminates between objects in the test set.

in all test cases and has a large bar graph element in the 2-state column of Figure 5. This suggests the mean separation of evaluation probabilities was very large and consistent between HMM 11 and all other HMMs for the given test set and also implies good performance against new test patterns. However Object 31, which was also correctly classified in all test cases, has a small bar element in the 3-state column of Figure 5. This means there is little separation between the correct and incorrect evaluation probabilities and this particular model may not generalize well to new patterns.

V. CONCLUSIONS AND FUTURE WORK

The promising results from this paper suggest that HMMs are a useful pattern classification scheme to discriminate between different acoustic returns generated by striking or other impulse excitation methods. Additionally, this approach worked well despite the restrictions of a small number of training sequences and the identical geometry of the targets. The two important areas this study did not address are: 1) intraclass feature variation and 2) signature aspect dependency. It is useful to know how well a given HMM can distinguish multiple objects of the same class from other object classes and what features have a low within-class variance. By using only one object to characterize a given object class, it is difficult to assert that another HMM trained on features gathered from multiple objects of the same class will have the same discrimination ability. It is well known that the acoustic signature of an object is aspect-dependent. Multiaspect HMM acoustic classification is addressed in [4] and [5] and involves training HMMs with acoustic returns taken over a sequence of different aspects. Addressing these two concerns for the case of acoustic impulse response classification is a topic of future work.

Finally, the time-frequency decomposition method discussed in this paper was not the only one explored in the work

TABLE I
FRACTION OF CORRECT CLASSIFICATION TO TOTAL TEST SEQUENCES

	Shell Type 1	Shell Type 2	Shell Type 3
Filler Type 1	$\frac{7}{7}$	$\frac{6}{7}$	$\frac{7}{7}$
Filler Type 2	$\frac{6}{7}$	$\frac{7}{7}$	$\frac{7}{7}$
Filler Type 3	$\frac{6}{7}$	$\frac{7}{7}$	$\frac{5}{7}$

of this research. The wave-based matching pursuits algorithm presented by McClure and Carin [7] was attempted with varying degrees of success for this application. If the impulse response can be calculated from a good understanding of the generating physics and more importantly these calculated effects can be observed in the striking experiments, then these customized basis decompositions show great promise. Future work will continue to explore the use of different basis decompositions that more compactly define the acoustic impulse response.

ACKNOWLEDGMENTS

This work was supported by the Naval Surface Warfare Center Panama City Independent Applied Research program funded by the Office of Naval Research.

REFERENCES

- [1] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, February 1989.
- [2] L. Rabiner, B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [3] A. Kundu, G. Chen, C. Persons, "Transient Sonar Signal Classification Using Hidden Markov Models and Neural Nets," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, January 1994.
- [4] P. Runkle, L. Carin, L. Couchman, J. Bucaro, T. Yoder, "Multiaspect Identification of Submerged Elastic Targets via Wave-Based Matching Pursuits and Hidden Markov Models," *J. Acoust. Soc. Am*, vol. 106, no. 2, August 1999.
- [5] Y. Dong, P. Runkle, L. Carin, "Markov Modeling of Transient Scattering and Its Application in Multi-Aspect Target Classification," *Proceedings ICASSP '02*, vol. 5, pp. 2841-2844, 7-11 May 2001.
- [6] Y. Linde, A. Buzo, R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-95, Jan 1980.
- [7] M. McClure, L. Carin, "Matching Pursuits with a Wave-Based Dictionary," *IEEE Transactions on Signal Processing*, vol. 45, no. 12, December 1997.